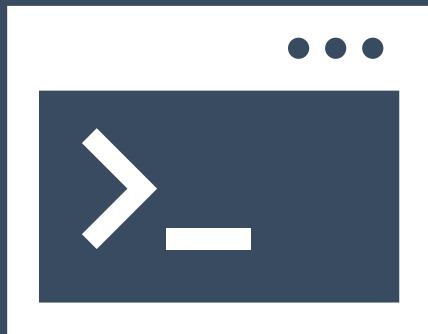


Differential expression analysis of Single Cell RNA-seq

<https://tinyurl.com/DGE-analysis-scRNAseq>



Harvard Chan Bioinformatics Core



Introductions!





Shannan Ho Sui
Director



Meeta Mistry
Associate Director



Lorena Pantano
*Director of Bioinformatics
Platform*



John Quackenbush
Faculty Advisor



Open Bhattarai



Heather Wick



Will Gammerdinger



Noor Sohail



Elizabeth
Partan



Alex Bartlett



Emma Berdan



James Billingsley



Zhu Zhuo



Maria Simoneau



Shannan Ho Sui
Director



Meeta Mistry
Associate Director



Lorena Pantano
*Director of Bioinformatics
Platform*



John Quackenbush
Faculty Advisor



Open Bhattarai



Heather Wick



Will Gammerdinger



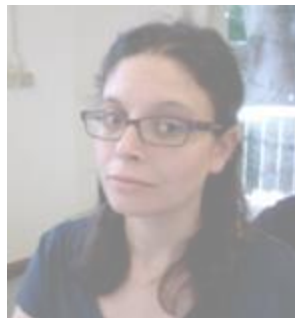
Noor Sohail



Elizabeth
Partan



Alex Bartlett



Emma Berdan



James Billingsley



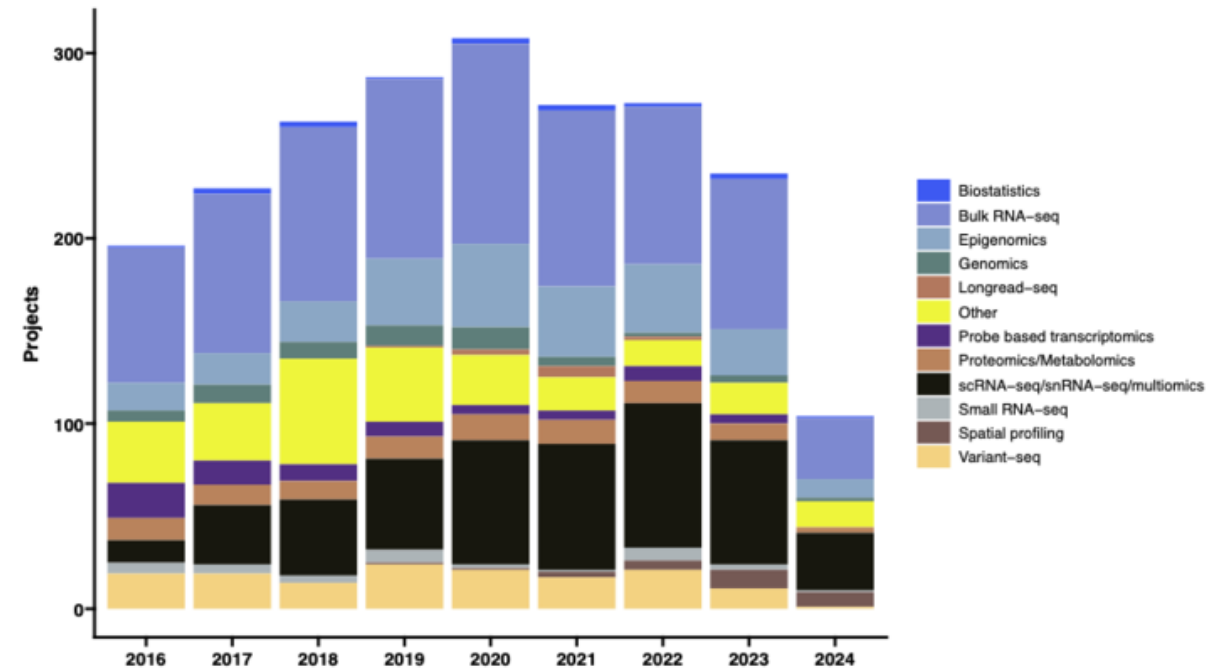
Zhu Zhuo



Maria Simoneau

Consulting

- ❖ Transcriptomics: Bulk, single cell, small RNA
- ❖ Epigenomics: ChIP-seq, CUT&RUN, ATAC-seq, DNA methylation
- ❖ Variant discovery: WGS, resequencing, exome-seq and CNV
- ❖ Multiomics integration
- ❖ Spatial biology
- ❖ Experimental design and grant support



Consulting

- ❖ Transcriptomics: Bulk, single cell, small RNA
- ❖ Epigenomics: ChIP-seq, CUT&RUN, ATAC-seq, DNA methylation
- ❖ Variant discovery: WGS, resequencing, exome-seq and CNV
- ❖ Multiomics integration
- ❖ Spatial biology
- ❖ Experimental design and grant support



HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH

NIEHS



THE HARVARD CLINICAL
AND TRANSLATIONAL
SCIENCE CENTER



HARVARD
MEDICAL SCHOOL

Training

- ❖ Hands-on workshops design to reflect best practices, reproducibility and an emphasis on experimental design
 - ❖ Basic Data Skills
 - ❖ Shell
 - ❖ R
 - ❖ Advanced Topics: Analysis of high-throughput sequencing data
 - ❖ Chromatin Biology
 - ❖ Bulk RNA-seq
 - ❖ Differential Gene Expression
 - ❖ scRNA-seq
 - ❖ Variant Calling
 - ❖ Current Topics in Bioinformatics

Training

- ❖ Hands-on workshops design to reflect best practices, reproducibility and an emphasis on experimental design
 - ❖ Basic Data Skills
 - ❖ Shell
 - ❖ R
 - ❖ Advanced Topics: Analysis of high-throughput sequencing data
 - ❖ Chromatin Biology
 - ❖ Bulk RNA-seq
 - ❖ Differential Gene Expression
 - ❖ scRNA-seq
 - ❖ Variant Calling
 - ❖ Current Topics in Bioinformatics

<https://bioinformatics.sph.harvard.edu/training>



HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH

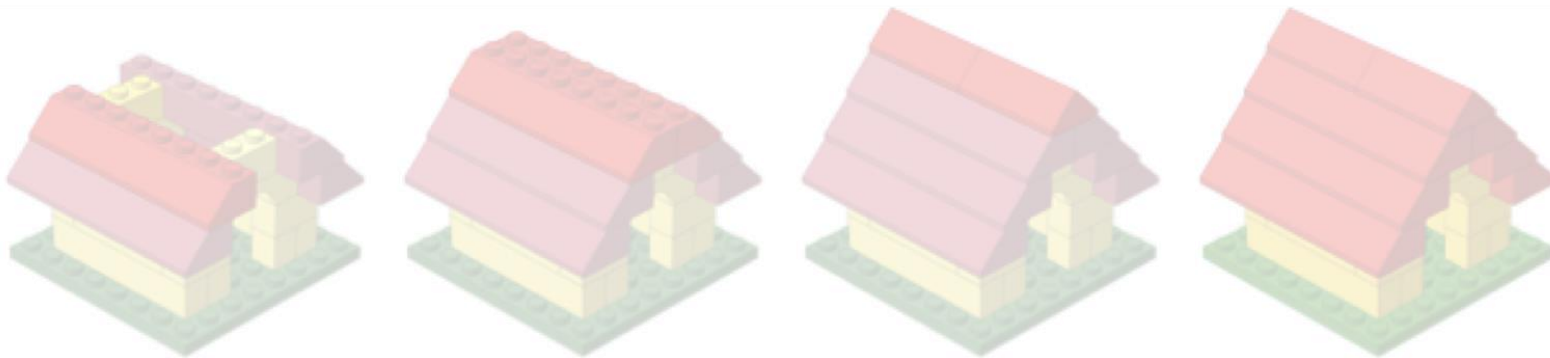
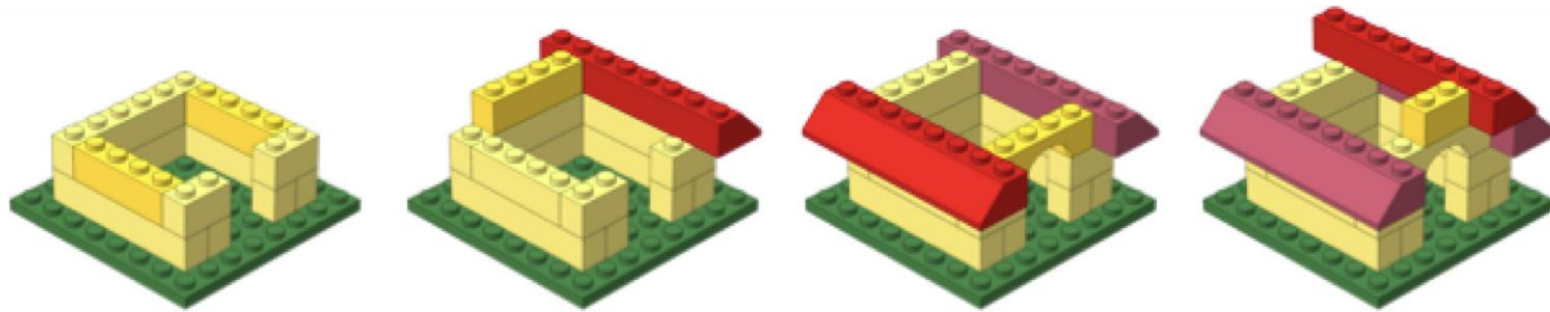
DF/HCC
DANA-FARBER / HARVARD CANCER CENTER



THE HARVARD CLINICAL
AND TRANSLATIONAL
SCIENCE CENTER

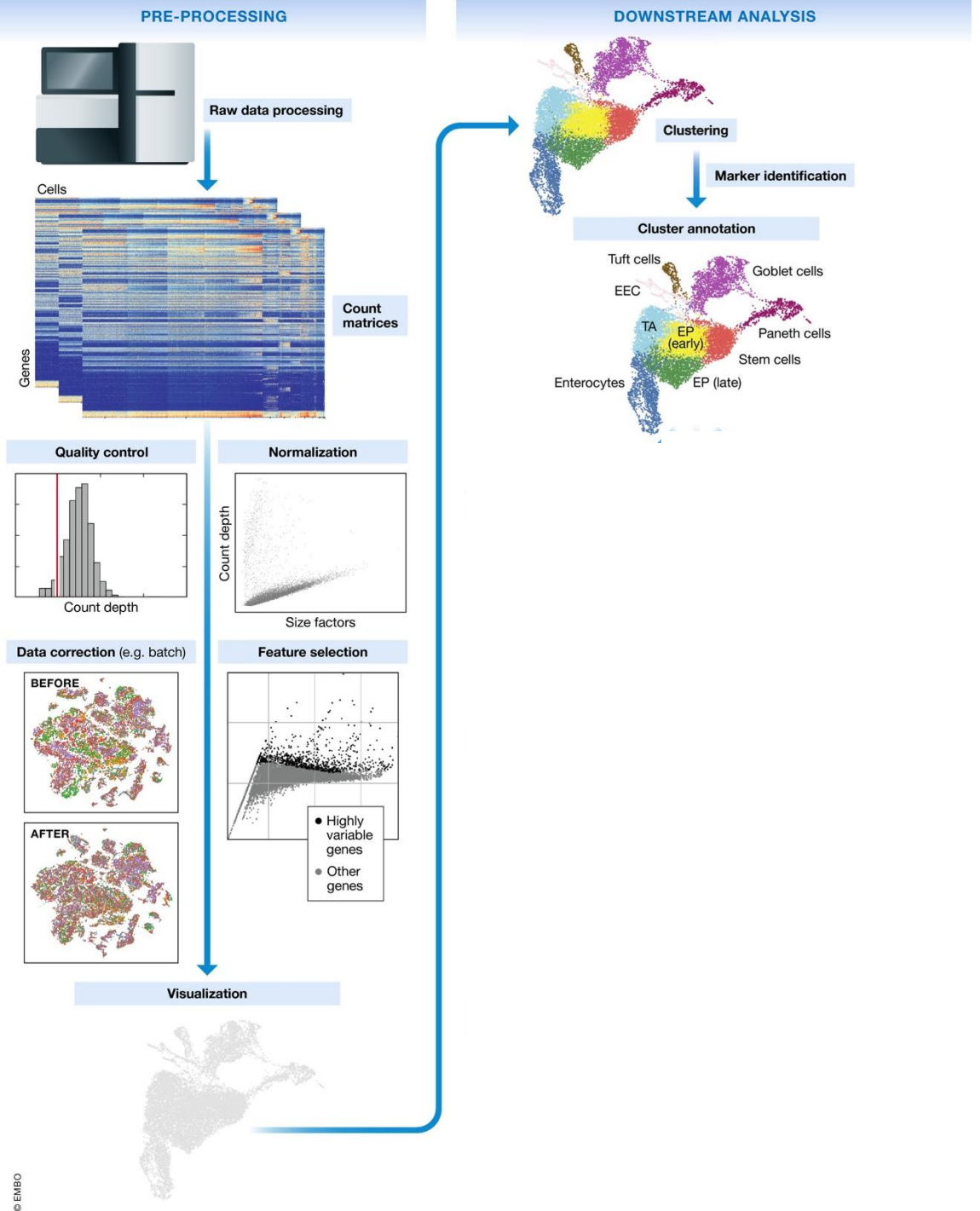


Workshop scope



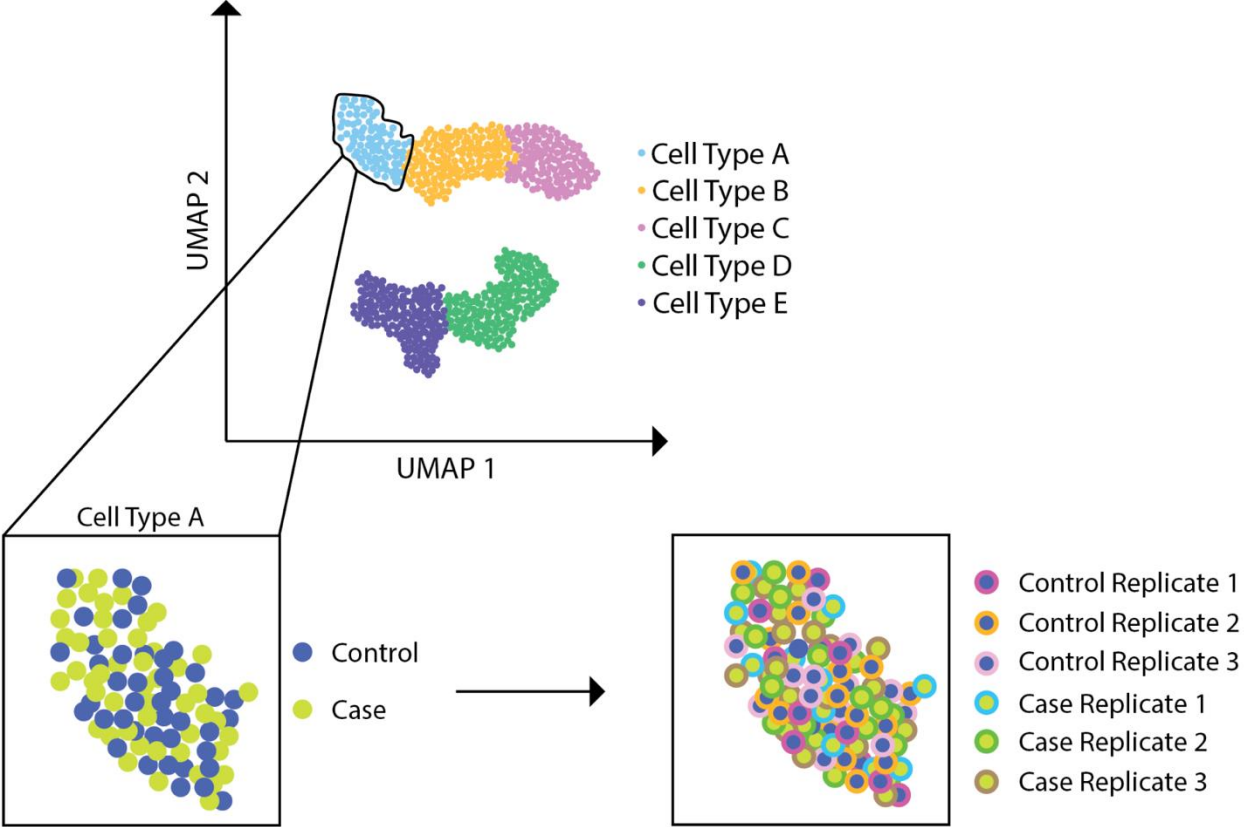
Bioinformatic Data Analysis

What is NOT covered in this workshop

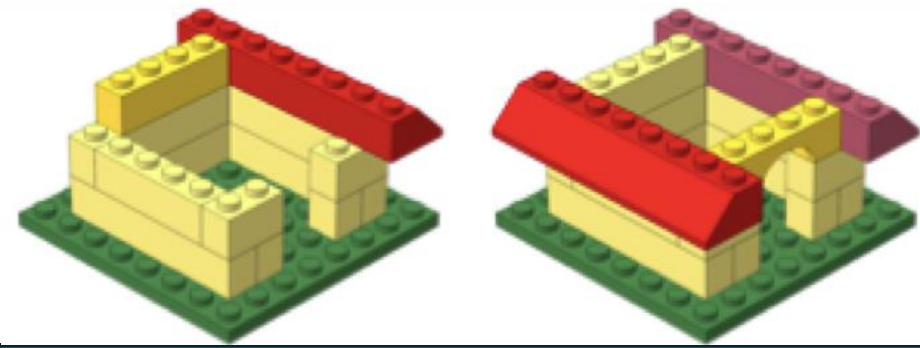


Luecken, MD and Theis, FJ. Current best practices in single-cell RNA-seq analysis: a tutorial, Mol Syst Biol 2019 (doi: <https://doi.org/10.15252/msb.20188746>)

What IS covered in this workshop



Workshop Scope



- ❖ Understanding considerations for when to use different DGE algorithms on scRNA-seq data
- ❖ Using FindMarkers to evaluate significantly DE genes
- ❖ Aggregating single cell expression data into a pseudobulk counts matrix to run a DESeq2 workflow
- ❖ Evaluating expression patterns of differentially expressed genes at the pseudobulk and single cell level
- ❖ Application of methods for evaluating differential proportions of cells between conditions

Logistics



Course schedule

Day 1

Time	Topic	Instructor
09:30 - 09:45	Workshop introduction	Noor
09:45 - 10:30	Project setup and data exploration	Noor
10:30 - 10:40	Break	
10:40 - 11:45	Differential expression analysis using <code>FindMarkers()</code>	Meeta
11:45 - 12:00	Overview of self-learning materials and homework submission	Meeta

Before the next class:

I. Please **study the contents** and **work through all the code** within the following lessons:

1. [Aggregating counts by celltype using pseudobulk approach](#)
Click here for a preview of this lesson
2. [DE analysis of pseudobulk data using DESeq2](#)
Click here for a preview of this lesson

<https://tinyurl.com/DGE-analysis-scRNAseq>

Course materials

- ❖ We continuously update our materials to reflect changes in the field/software



Approximate time: 75 minutes

Learning Objectives:

- Evaluate differential gene expression between conditions using a Wilcoxon rank sum test
- Create visualizations for differentially expressed genes
- Discuss other statistical tests for differential expression analysis

Differential expression between conditions using `FindMarkers()`

In our current UMAP, we have merged samples across the different conditions and used integration to align cells of the same celltype across samples. Now, what if we were interested in a particular celltype and **understanding how gene expression changes across the different conditions?**

<https://tinyurl.com/DGE-analysis-scRNAseq>

Single Screen & 3 Windows

The image displays a single screen with three overlapping windows. The top-left window is a Zoom meeting interface showing three participants: Mary Piper, Troubleshooter (Radhika), and Jihe Liu. The top-right window is a web browser displaying a page with a 'View on GitHub' button. The bottom window is the RStudio IDE, showing R code being executed in the console. The code includes comments for assignment operators, functions, and rounding functions. The console output shows the results of these operations, including the value of x (3) and the result of round(3.14159) (3).

```
1 # Assignment operator
2 x <- 3
3
4 # Functions
5 getwd()
6
7 sqrt(81)
8
9 round(3.14159)
10 ?round
11
```

Environment

Variable	Value
x	3

Values

Variable	Value
number	15
x	5
y	10

Console

```
> x <- 3
> # Functions
> getwd()
[1] "/Users/mariyaper/Desktop/R-testing"
> sqrt(81)
[1] 9
> round(3.14159)
[1] 3
> ?round
>
```

Rounding of Numbers

Description

ceiling takes a single numeric argument x and returns a numeric vector containing the smallest integers not less than the corresponding elements of x.

floor takes a single numeric argument x and returns a numeric vector containing the largest integers not greater than the corresponding elements of x.

trunc takes a single numeric argument x and returns a numeric vector containing the integers formed by truncating the values in x toward 0.

round rounds the values in its first argument to the specified number of decimal places (default 0). See 'Details' about "round to even" when rounding off a 5.

signif rounds the values in its first argument to the specified number of significant digits.

Usage

```
ceiling(x)
floor(x)
trunc(x, ...)
```

Single Screen & 3 Windows

Zoom

Our Recommendation

```
# Assignment operator
x <- 3

# Functions
getwd()
sqrt(81)
round(3.14159)
?round
```

```
> x <- 3
> # Functions
> getwd()
[1] "/Users/mariyaper/Desktop/R-testing"
> sqrt(81)
[1] 9
> round(3.14159)
[1] 3
> ?round
```

Environment History Connections

Global Environment -

Values

Variable	Value
x	3

Files Plots Packages Help Viewer

R: Rounding of Numbers - Find in Topic

Round (base) R Documentation

Rounding of Numbers

Description

`ceiling` takes a single numeric argument `x` and returns a numeric vector containing the smallest integers not less than the corresponding elements of `x`.

`floor` takes a single numeric argument `x` and returns a numeric vector containing the largest integers not greater than the corresponding elements of `x`.

`trunc` takes a single numeric argument `x` and returns a numeric vector containing the integers formed by truncating the values in `x` toward 0.

`round` rounds the values in its first argument to the specified number of decimal places (default 0). See 'Details' about "round to even" when rounding off a 5.

`signif` rounds the values in its first argument to the specified number of significant digits.

Usage

```
ceiling(x)
floor(x)
trunc(x, ...)
```


Single Screen & 3 Windows

The screenshot displays a Zoom meeting interface. At the top, there are three video thumbnails for participants: Mary Piper, Troubleshooter (Radhika), and Jihe Liu. Below the thumbnails is a toolbar with icons for muting, video, and sharing. The main content area shows a shared RStudio window. The RStudio window has a dark theme and is divided into several panes. On the left is the source editor with R code. At the bottom is the console showing the execution of the code. On the right is a pane titled 'Web Browser' which displays the help page for the 'round' function in R. A green box highlights the 'Web Browser' pane and the console output.

```
483  
484  
485 getwd()  
486  
487 # square root function  
488 sqrt(81)  
489  
490 # round function  
491 round(3.14159)  
492 ?round  
493  
494  
495
```

```
1 # Assignment operator  
2 x <- 3  
3  
4 # Functions  
5 getwd()  
6  
7 sqrt(81)  
8  
9 round(3.14159)  
10 ?round  
11
```

```
> x <- 3  
> # Functions  
> getwd()  
[1] "/Users/marypiper/Desktop/R-testing"  
> sqrt(81)  
[1] 9  
> round(3.14159)  
[1] 3  
> ?round  
>
```

Web Browser

Rounding of Numbers

Description

integers not less than the corresponding elements of x.

floor: takes a single numeric argument x and returns a numeric vector containing the largest integers not greater than the corresponding elements of x.

trunc: takes a single numeric argument x and returns a numeric vector containing the integers formed by truncating the values in x toward 0.

round rounds the values in its first argument to the specified number of decimal places (default 0). See 'Details' about "round to even" when rounding off a 5.

signif rounds the values in its first argument to the specified number of significant digits.

Usage

```
ceiling(x)  
floor(x)  
trunc(x, ...)
```

*Our
Recommendation*

Single Screen & 3 Windows

The image displays a Zoom meeting interface at the top, showing three participants: Mary Piper, Troubleshooter (Radhika), and Jihe Liu. Below the Zoom window, there are three overlapping windows. The leftmost window is a terminal window showing R code and its output. The middle window is the RStudio IDE, displaying R code for assignment operators, functions, and rounding. The rightmost window is a browser window showing a GitHub page for 'Rounding of Numbers'.

```
483  
484  
485 getwd()  
486  
487 # square root function  
488 sqrt(81)  
489  
490 # round function  
491 round(3.14159)  
492 ?round  
493  
494  
495
```

```
1 # Assignment operator  
2 x <- 3  
3  
4 # Functions  
5 getwd()  
6  
7 sqrt(81)  
8  
9 round(3.14159)  
10 ?round  
11
```

```
> # round function  
> round(3.14159)  
[1] 3  
> ?round  
>
```

```
> x <- 3  
> # Functions  
> getwd()  
[1] "/Users/mariyaper/Desktop/R-testing"  
> sqrt(81)  
[1] 9  
> round(3.14159)  
[1] 3  
> ?round  
>
```

Our Recommendation

R Studio

Single Screen & 3 Windows

Zoom

Web Browser

R Studio

```
> x <- 3
> # Functions
> getwd()
[1] "/Users/mariyaper/Desktop/R-testing"
> sqrt(81)
[1] 9
> round(3.14159)
[1] 3
> ?round
>
```

Participants (3)

- Mariyaper (Co-host, me)
- Jihe Liu (Host)
- Troubleshooter (Radhika) (Co-host)

Values

number	15
x	5
y	10

Environment

Global Environment	
Values	
x	3

Files | Plots | Packages | Help | Viewer

R: Rounding of Numbers

Round (base)

Rounding of Numbers

Description

integers not less than the corresponding elements of x.

floor: takes a single numeric argument x and returns a numeric vector containing the largest integers not greater than the corresponding elements of x.

trunc: takes a single numeric argument x and returns a numeric vector containing the integers formed by truncating the values in x toward 0.

round: rounds the values in its first argument to the specified number of decimal places (default 0). See 'Details' about "round to even" when rounding off a 5.

signif: rounds the values in its first argument to the specified number of significant digits.

Usage

```
ceiling(x)
floor(x)
trunc(x, ...)
```

Our Recommendation

R Studio

Course participation

- ❖ Mandatory review of self-learning lessons and assignments
- ❖ Attendance required for all classes
- ❖ Your questions and active participation drive learning
- ❖ **We look forward to all of your questions!**



Course participation

- ❖ At-home lessons and exercises after each session
- ❖ Cover material not previously discussed
- ❖ Provides us feedback to help pace the course appropriately
- ❖ 3-5 hours to complete
- ❖ Homework load is heavier in the beginning of this workshop series and tapers off

Using AI for Assignments



❖ Do

- ❖ Try to resolve error messages with it
- ❖ Test code written by AI on a dataset where you have expected results
- ❖ Take the time to review the generated code line-by-line

❖ Don't

- ❖ Implement it in replacement to learning
- ❖ Write code that you don't understand
- ❖ Assume the output from an AI process is correct

Odds & Ends

- ❖ Quit/minimize all applications that are not required for class
- ❖ Name tags
- ❖ Post-its
 - ❖  green - I am all set
 - ❖  red - I need time/help
- ❖ Phones on vibrate/silent
- ❖ Bathrooms

Contact Us

- ❖ *HBC training team:* hbctraining@hsph.harvard.edu
- ❖ *HBC consulting:* bioinformatics@hsph.harvard.edu